SUMMARY

KNOWLEDGE REPRESENTATIONS AND SYMBOLS IN THE MIND.

The central issue in this dissertation is the development and the application of an instrument with which to compare several theories of mental representation which are at present current in cognitive psychology. The main reason for developing such an instrument lies in the confusion and the lack of consistency which exists at the present time in relation to central notions in cognitive psychology. The instrument developed consists of five criteria which are levels of description (c1), morphological criterion (c2), n-place predicate (c3), direction of reference (c4) and characteristics of symbols (c5).

In the first part of this study the problem will be formulated (chapter 1) and the instrument developed (chapter 2). In the second part six cognitive theories will be described and subsequently evaluated in the light of the criteria. The six selected include Marr's theory of perception (chapter 3), Kosslyn's theory of mental imagery (chapter 4), Tulving's theory of episodic memory (chapter 5), Anderson's ACT*-theory (chapter 6), Kintsch's theory of propositional representation (chapter 7) and finally Schank & Abelson's theory of scripts and themes (chapter 8). In the third part, an overview of the results of the application of the criteria will be given (chapter 9) and finally, several conclusions (chapter 10) will be derived concerning necessary conditions for a computational theory of mind.

Part 1: The problem and the instrument.

In chapter 1 the background, the assumptions and the central notions of cognitive psychology are described. Cognitive psychology focusses on the study of memory, language, perception, problem solving and thinking. The conceptual impact of cognitive psychology is deeply connected with the development of the (digital) computer. The underlying connection is the idea that the human mind and the computer are both instances of information processing systems which share characteristics such as the manipulation of symbols, the conditional structure of information and the exchangeable representational format of data and processes.

The central concepts in cognitive psychology are the concepts of symbol manipulation, of mental representation and of computation. "Symbol manipulation" implies that the entities on which the processing takes place can be viewed as symbols. Symbols, in their turn, can be conceived as the constituent parts of "representations". "Computation" is manifested in the idea that thinking consists of the processing, storage and retrieval of information.

Summary.

The problem with these concepts is twofold. In the first place, these concepts are very complex and have ambiguous meanings and in the second place, which is partly a consequence of the first, there is still no unified theory of cognition. Many models and theories have been formulated in cognitive psychology, but these do not fit together effortlessly and indeed often contradict one another.

A comparison between theories can be carried out in two ways. The first way is by engaging in empirical research. This is what I call an internal comparison. The second alternative is to compare central concepts in theories by means of criteria. This is called an external comparison and is the method selected for this thesis.

The principal problem that will be highlighted in this dissertation is as follows: given the concepts of symbol system, representation and computation, how is it possible to compare various cognitive theories and what are the results of these comparisons?

Newell & Simon have been the most outspoken on the symbolical and representational nature of human and artificial intelligence. According to these authors, humans and computers are examples or realizations of physical symbol systems. Symbol systems are characterized by the fact that they behave flexibly and "symbolically" and that they can be seen as machines with effectively computable procedures. The prefix "physical" does not imply a reduction to physics, but applies to a physical realization and the boundaries within which such a "machine" operates. The conceptual framework of symbol systems is the basis for the cognitive theory of Newell & Simon, within which all intelligent action, such as decision making, reasoning, and thinking are considered as forms of problem solving.

In the view of Fodor (1981), the research strategy in cognitive psychology can be called "mental representationalism". Reasoning and problem solving in the human system consists of the processing of (mental) representations. The problem with this statement is that "representation" can be used ambiguously as, for example, in propositional representation, pictorial representation, procedural representation, declarative representation, representational content and processes of representation. Many controversies in cognitive theory suppose some notion and interpretation of "representation". By comparing the views of Anderson and Fodor in relation to the unity of the mind it is made clear that the notion of representation is used a) in the sense of a set of cognitive entities, b) in the sense of computation or procedure and c) in the sense of a model of the (physical) world. This last sense focusses on the aspect of depiction or the mirroring of a (mental) representation.

Concerning the notion of computation, it is not the notion itself but its assumption of functionalism which is most interesting in this respect. The idea is that computation as the processing of information works independently of

the physical structure of a system. According to Block (1981) this starting-point is called "computation-representation" functionalism. This variant of functionalism describes the functioning of the human mind in terms of a computer program. Psychological states are internal states, that is to say mental representations on which computations are executed. The emphasis in functionalism is on the definition, the functioning and the coherence of the components and the sub-components of the system.

The criteria that have been used to characterize and to evaluate the six cognitive theories have been derived from the notion of computation (criterion 1), representation (criteria 2, 3, 4) and symbol system (criterion 5). Furthermore, every criterion consists of several labels, what I have called "valences". By this I mean that, with respect to a criterion, every cognitive theory gets a value, which indicates the status of that theory.

In chapter 2 the instrument in the form of five criteria is discussed. Criterion 1 (c1) is about different levels of description and distinguishes a physical, a functional and an intentional level in order to predict and to explain the behaviour of complex systems (Dennett, 1978). In the physical stance one tries to use the laws of nature. In the functional stance the only important aspect is the proper functioning of a system. Neither the physical realization of the processes and mechanisms nor physical structure is relevant. What counts in a functional description is the definition of the components and the relation between the components of the system. The intentional stance is necessary if the system is very complex, so that we can no longer use the physical and functional descriptions. Instead, intentions, wishes and expectations are ascribed to the system. But we must remember that in so doing we are trying to explain rationality in terms of rationality.

With regard to the physical, functional and intentional stance the ideal level of description for cognitive psychology is the functional stance. It is therefore very important to separate the levels of description. Valence 1 implies that a cognitive theory or model is *functional* in its description. Valence 2 indicates that the *separation* between the *physical* and *functional* level of description is *neglected*, whereas valence 3 implies that the *separation* between the *functional* and *intentional* level is *neglected*. Valence 4 indicates that in the theory all *necessary* distinctions of levels are *disregarded*.

The second criterion (c2) - the morphological criterion - relates to the various interpretations of the notion of representation that can be distinguished. This can be illustrated by referring to the following list of overlapping "synonyms": on the one hand, description, model, symbol set, picture and sign-system; on the other hand, analogy, depiction, reference, correspondence and denotation. In the former group the accent is on the domain or the domains in the representation, whereas in the latter group the relation between two domains is emphasized.

Summary.

With regard to the nature of the domains, Palmer (1978), has formulated five requirements for a representation. In the first place it should be clear what the represented domain is and secondly what the basic elements or entities in this domain are. In the third place it should be clear what the representing domain is and fourthly what the basic elements in this domain are. In the fifth place one should establish the correspondences, that is to say the relation, between the domains.

Two interpretations can be discerned concerning the representation *relation* as a connection between domains. The first refers to a relation between representing domains, which I will describe as being the equivalence of domains, and the second refers to a relation between a representing and a represented domain. This is called the correspondence or the similarity of domains.

In cognitive psychology it is often suggested that representation as a relation presupposes some sort of similarity of domains. A pictorial representation, for example, is supposed to be similar to what it is actually a representation of. This position seems extremely unlikely and the position which will be adopted in this dissertation is that one of the main problems that theories of mental representation face, is the so called a priori necessity of a natural resemblance between a represented and a representing domain.

Valences of the morphological criterion are 1) that the *entities* in the representing domain are *defined*. Valence 2) implies that in a cognitive theory representation is conceived as a *symbol set* and valence 3) that representation is conceived as a *process* or *procedure*. Valence 4) indicates that representation is used in the sense of *depiction*. Valence 5) is about the question of whether *resemblance* of domains is or is not a necessary condition for representation.

The third criterion (c3) is the criterion of n-place predicates. N-place predicate means that the representation as a relation takes n predicates. Working with two domains gives the impression that representation is always a two-place predicate. Nevertheless representation can also be used as a one-place or a three-place predicate.

Representation as a one-place predicate is equivalent to representation in the sense of symbol set. In such a case there is no indication of reference or mirroring. What representation does, is that it classifies or categorizes.

The accepted view in cognitive psychology is that a representation is a two-place predicate, that is to say y is a representation *of* x (2a), for example a mental image of a tree. The position is more complicated with two-place predicates, since it is possible to say that y is a representation *for* z (2b), for example, this is an image for or in my consciousness. In these two cases x is standing for the external world, y is the mental system, and z is "consciousness" or "person".

If representation is a three-place predicate, then y is a representation *of* x *for* z (3). The problem with this point of view is that it is almost impossible to determine y (mental system) independently of z (consciousness). The four valences of this criterion correspond to the numbers given between brackets.

The fourth criterion (c4) concerns a different aspect of the notion of representation in that it relates to the direction of reference. For example, domain A can refer to domain B and vice versa. Here it is possible to discern literal reference and figurative reference. The expression "John kicks the ball" has a literal reference, whereas "John is a fox" has a metaphorically one. In the same way, a picture of my house refers literally to my house, whereas Picasso's "Guernica" refers metaphorically to the cruelty of war. Complex reference is a combination of literal and non-literal reference.

Denotation differs from non-denotation in the direction of reference. In denotation the direction is from symbol set to object or state of affairs, whereas in non-denotation an object literally or metaphorically refers to something which could be an object or a symbol (set), but which in any case literally or metaphorically possesses the property by which it refers to. If an object literally refers and possesses the property it is referring to in directing the attention in some respect, it is called exemplification. For example, a tailor's swatch directs the attention to the colour or the texture and not to the size of what it is referring to. This aspect of attention makes exemplification different from instantiation. When the referring is metaphorically, it is called expression, for example in a musical overture which is "tragic".

Valences of this criterion of direction of reference are: 1) *denotation*, 2) *exemplification* or 3) *expression*.

The fifth criterion (c5) relates to the characteristics of symbols. Symbol sets can be distinguished by examining their syntactical and semantical aspects. If a symbol set is called a notation, then it is defined by two syntactical and three semantical requirements. The primary function of a notation, as the most rigidly formulated symbol set, is "authoritative identification". The syntactical requirements are 1) disjointness, which means that symbols might be replica's of one another, and 2) finite differentiation, which means that it should be possible to decide whether a sign belong to the one or to the other symbol. The semantical requirements are 1) unambiguity of the symbols, 2) disjointness, which means that the symbols should not be redundant, and 3) finite differentiation which means that a symbol should apply to at least one reference-class.

Valence 1) of this criterion indicates that none of the requirements is satisfied; we just have a *symbol set*, valence 2) that only the syntactical requirements are satisfied and in this case we speak of a *notational scheme*. Valence 3) indicates that both the syntactical and semantical requirements are fulfilled in which case we speak of a *notation* or a *notational system*.

Summary.

Part 2: The material: six theories of mental representation.

In this part, six theories of mental representation - a fair sample of the theories that can be found in cognitive science - are described and evaluated.

In chapter 3, Marr's theory of perception is discussed. According to Marr, three perceptual levels should be discerned in the human system. Each of these levels consists of basic input and output elements and of algorithms which describe the transformation of information from input to output. The first level is called the primary sketch which gives the changes of intensity and the geometrical organization of the surfaces. The output-representation of this level is the input-representation for the next stage, which gives the 2,5-D sketch. This sketch contains the combination of several modular components of the perceptual system. The 2,5-D sketch is perceiver-oriented. The next stage is the 3-D sketch which is entirely object-oriented. The 3-D sketch gives the information we normally have when we see horses, mountains and cars and it works with volumetric entities as primitives and is closely connected with the higher cognitive processes.

This description is followed by an evaluation based on the five criteria. In Marr's theory the separation in a functional and a physiological level of description is neglected (c1). Furthermore, concerning the morphological criterion (c2), Marr's theory uses representation in the sense of symbol set and depiction, whereas the entities in the represented domain are clearly defined. In the criterion of n-place predicates (c3), Marr takes the position that y is a representation of x, that is to say representation is a 2a-place predicate. In his theory the direction of reference (c4) is from symbol sets to states of affairs which means that it is mainly denotation. With respect to the last criterion (c5), the symbols in Marr's theory constitute a notational scheme and not a notational system.

In chapter 4, Kosslyn's theory of mental images is portrayed. Kosslyn uses the metaphor of the cathode ray tube in order to describe and to explain the nature and processes of mental imagery. The surface level gives the vivid impression of "seeing" something with a mind's eye and which consists of quasi-pictures in a visual buffer. These quasi-pictures are constructed from two lower levels of representations, a level of so called literal representation which has a skeletal structure and a deeper level in long term memory in which the representations are proposition-like. Besides all these representations, there are many processes within as well as between the levels of representation.

Kosslyn's theory is mainly formulated in functional terms (c1), but physiological and intentional levels of description are frequently used implicitly. Furthermore, it is questionable whether the entities in the represented domains are clearly formulated (c2). This makes it unclear whether he uses representation in the sense of symbol set. It could be stated that he uses representation

in the sense of procedure and in the sense of depiction. There is no indication of resemblance between an inner and an outer domain from which it follows that representation is used as a 2b-place predicate (c3). Furthermore, his mentioning of different levels of representation makes it clear that each level is an expression of another level (c4). Although Kosslyn tries to defend the position that representation is without symbol sets, it is clear that his model only functions if one assumes that there are structures which constitute representations. The unwillingness to define symbol structures in a more precise way, makes it clear that the symbol sets in mental imagery are neither notational schemes nor notational systems (c5).

In chapter 5, Tulving's theory of episodic memory is discussed. According to Tulving, the research on human memory is dominated by semantical and propositional structures. This, he argues, is suspect, since it does not account for the fact that people remember episodes and particular situations. For this reason he proposes an episodic memory system in contrast to a semantic memory system. The model that Tulving has formulated for the episodic system is called the General Abstract Processing System (GAPS). This model consists of two sections, storage and retrieval. In the first part of the model, after encoding, the information is stored in what Tulving calls an original engram, whereas in the second part, the information is retrieved as a function of the matching of the retrieval cue and the encoded and recoded information in episodic memory. The storage and retrieval are independent of semantical structures.

In Tulving's theory of episodic memory descriptions of the functional, physiological and intentional level are all intermingled (c1). Furthermore, it is not clear whether he defines entities in the represented domain, which means that he refuses to speak of representations in the sense of symbol sets, although representations in Tulving's theory are used in the sense of procedure and depiction (c2). With regard to representation as an n-place predicate Tulving's theory has a two-sided orientation. Since there is a similarity between episodes outside and inside the human system, Tulving argues that representation is a 3-place predicate, that is to say some internal entity is a representation of an episode for a person's consciousness. A further point is that representation is also a 2b-place predicate (c3). The direction of reference according to Tulving is mainly from internal system to episodes in the real world, which means that there is denotation (c4). Tulving refuses to accept general symbol structures, so his theory is neither a notational scheme nor a notational system. Doubts are raised as to whether GAPS has general symbol structures in any case (c5).

Chapter 6 is about Anderson's theory of cognition which is called the theory of the Adaptive Control of Thought (ACT*). ACT* is semantically or propositionally orientated and is based mainly on the theory of production

Summary.

systems. A production rule consists of a condition-action pair in which, when a condition is satisfied, an action is carried out. According to Anderson, a cognitive system consists of a declarative memory, a production memory and a working memory. The way in which the system operates is by the spreading of activation through all sorts of nodes and paths. Furthermore, Anderson distinguishes three kinds of mental representations: propositional representations, pictorial representations and temporal string representations.

Anderson's ACT*-theory is mainly formulated in functional terms. Nevertheless he makes references towards physiological levels of description and in ACT* there are also indications of intentional terms (c1). Concerning the morphological criterion, Anderson tries to define entities for all forms of mental representation and, over and above, he uses the concept of representation in the sense of procedure and depiction (c2). The aspect of different forms of mental representation also has implications for representation as n-place predicate. In Anderson's theory, representation is on the one hand a 2a-place predicate and on the other a 2b-place predicate (c3). This many-sided interpretation of representation means that the direction of reference is a denotation and an exemplification (or expression) (c4). Anderson certainly tries to define symbol sets in ACT* and, in respect to propositional representations, tries to define a notational scheme. In a way a production can be seen as such. In ACT* no notational system is defined for any of the forms of mental representation.

In chapter 7, Kintsch's theory of propositional representations is discussed. Kintsch has formulated a theory of memory in order to explain how people understand text fragments. His theory has two central issues. In the first place Kintsch presupposes that knowledge and meaning in memory is represented in the form of propositions. The structure of the representations is a combination of a case grammar and predicate-argument pattern. In the second place, Kintsch assumes that the information in a text is partly explicit and partly implicit or deducible. The latter information is stored in the semantical memory system in the same propositional format as the explicit information. Besides a semantic memory, Kintsch also mentions an episodic system and both are closely connected.

Kintsch's theory is formulated mainly in functional terms, but implicitly there are some indications of physiological concepts and sometimes he also uses intentional terms (c1). Using the morphological criterion, it is clear that Kintsch tries to define the entities of representation which means that he uses representation in the sense of symbol set. He does not use representation in the sense of procedure and depiction. Similarity of domains is not a requirement in his theory (c2). Furthermore, in his theory, representation is mainly a 2b-place predicate (c3) and this means that the direction of reference is such that it is a combination of exemplification and expression. It is not denotation

(c4). Concerning the aspects of symbols, the entities in memory which Kintsch uses are such that they constitute a notational scheme, but not a notational system (c5).

Chapter 8 discusses Schank & Abelson's theory of scripts. This theory of scripts has been shown to be fundamental for much research in the field of artificial intelligence. The most well known example used by Schank & Abelson portrays the story of a visit to a Mcdonald's restaurant. Suppose someone goes to Mcdonald's, orders some food, pays the bill, gives a big tip and leaves the restaurant. What could one infer from this event? The point is that, according to Schank & Abelson, people have scenario's in their head by which they understand, plan and act. People normally have more or less complex chains of causes in their head which could be called scripts, plans, goals and themes. The model Schank & Abelson propose consists of two levels. The lower level is called the level of conceptual dependency, whereas the level above is called the level of the knowledge structure.

Schank & Abelson do not use physiological terms. The level of description they use is mainly a functional one, but for several aspects of their theory intentional terms are also applied which means that rationality of the system is presupposed instead of explained (c1). Furthermore, they use the notion of representation in the sense of symbol set, procedure and depiction and by using conceptual dependency and knowledge structures try to define the elements in the representing domain (c2). Since they use representation in the sense of depiction and symbol set, representation is applied on the one hand as a 2a-place predicate, whereas on the other hand, representation is applied as a 2b-place predication (c3). With respect to the criterion of the direction of reference (c4), this relation in Schank & Abelson's theory on the one hand forms a denotation, on the other hand an exemplification (or expression). The entities Schank & Abelson define in the conceptual dependency and the knowledge structure are such that they form a notational scheme, but not a notational system (c5).

Part 3: Conclusion and epilogue.

Chapter 9 contains three surveys of the criteria. The first summarizes the valences of the levels of description, the second the valences of the morphological criterion, the criterion of the n-place predicates and the criterion of the direction of reference and the third the valences concerning the requirements of symbol sets.

Chapter 10 discusses the results of the application of the criteria and is divided into three parts. The first part is about the confusion concerning the notion of representation, the second examines the characteristics of symbols

and the mental structure and the final part considers the necessary conditions for a computational theory of mind.

The confusion around the notion of representation used in cognitive psychology can be resolved by clearly distinguishing representation in the sense of symbol set, of procedure and of depiction. This separation also makes it clear in which respects propositional representations differ from, for example, analog representations or semantical representations. The instrument of the five criteria makes it easier to articulate the differences between the various forms of representation.

The conclusion suggests that when the syntactical and semantical requirements of symbols are considered, none of the cognitive theories that have been discussed succeed in formulating a notational system or notation. Nevertheless, it is argued that because of the fact that authoritative identification of structures is at the centre of theories of symbol manipulation, at least a notational scheme is necessary for any theory of mental representation. Since a proposition is discrete and well formed which together are constituents of a notational scheme, it is therefore easy to understand why propositional representations have such a prominent place in recent cognitive psychology.

The third and last part of the conclusion discusses the necessary conditions for a computational theory of mind. The conditions follow from the valences of the criteria, that is to say that, because of the specification of the valences, it is possible to isolate the "ideal" valences for a computational theory of mind.

In the first place such a theory should be formulated on a functional level of description and it should be made clear how the loans that are taken out of rationality could be settled.

In the second place, the entities in the representing domain should be defined, that is to say, they should at least form a notational scheme.

In the third place, representation as a three-place predicate should be avoided, since it involves a circular argument. Furthermore, the whole idea of a (natural) similarity between a represented and a representing domain should be abandoned. Resemblance of domains is a consequence of the choice for a representation and not the reason for it. Cognitive psychology should stop worrying about the represented domain, that is to say about the external world.

In the fourth place, a differentiation between representation in the sense of symbol set, procedure and depiction is recommended.

The conclusion which is drawn from the above mentioned aspects is that cognitive psychology does not yet contain a computational theory of mind without shortcomings. Nevertheless, the foundations of cognitive science are already present, the walls are being erected and positive steps are being taken towards establishing the final edifice (Newell, 1987).